DOCUMENT RESUME

ED 469 762 TM 034 549

AUTHOR	Fraas, John W.; Drushal, J. Michael; Graham, Jeff
TITLE	Expressing Logistic Regression Coefficients as Changes in Initial Probability Values: Useful Information for Practitioners.
PUB DATE	2002-10-17
NOTE	34p.; Paper presented at the Annual Meeting of the Mid-Western Educational Research Association (Columbus, Ohio, October 16-19, 2002).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	Computer Software; *Probability; *Regression (Statistics); Test Interpretation

ABSTRACT

This paper presents a method designed to assist practitioners in the interpretation of the practical significance of a statistically significant logistic regression coefficient is presented. To avoid the interpretation problems encountered when using the traditionally reported change in either the log odds or odds values, this method centers the interpretation on the change in the probability value of the event occurring associated with a given change in the independent variable for a range of initial probability values. To assist in the implementation of this method, a computer program, which is used in conjunction with the Microsoft (registered) Excel software, was designed to compute these probability values. An appendix contains the commands used in conjunction with the Excel software. (Contains 4 figures and 27 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 469 762

Running head: EXPRESSING LOGISTIC REGRESSION COEFFICIENTS

Expressing Logistic Regression Coefficients as Changes in Initial Probability Values: Useful Information for Practitioners

John W. Fraas and J. Michael Drushal

Ashland University

Jeff Graham

Medina City Schools

Paper presented at the Annual Meeting of the

Mid-Western Educational Research Association

October 17, 2002

TM034549

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

J. Fraas

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

Abstract

A method designed to assist practitioners in the interpretation of the practical significance of a statistically significant logistic regression coefficient is presented. To avoid the interpretation problems encountered when using the traditionally reported change in either the log odds or odds values, this method centers the interpretation on the change in the probability value of the event occurring associated with a given change in the independent variable for a range of initial probability values. To assist in the implementation of this method, a computer program, which is used in conjunction with the Microsoft® Excel software, was designed to compute these probability values.

Expressing Logistic Regression Coefficients as Changes in Initial Probability Values: Useful Information for Practitioners

It is not uncommon for program evaluators and researchers to encounter a variety of situations in which the dependent variable of interest is dichotomous (i.e., the variable consists of two categories). For example, a study by Graham (2002) investigated whether specific teacher background variables could discriminate between groups of teachers who were classified as belonging to either the early or late stages of concern regarding the implementation of an educational change in their schools. Schreiber (2002) attempted to determine if certain student background and academic factors could be used to identify whether the students scored above or below the international mean on an advanced mathematical examination. In a study by McCoach and Siegle (2001) attitude variables were analyzed to determine if they could accurately identify whether a student was a gifted achiever or a gifted underachiever. And Hendel (2001) examined if the students' participation in a first-year seminar and various academic and background variables were associated with whether they did or did not return for their second year of college.

When such dependent variables are encountered, it is not uncommon for researchers and program evaluators to use logistic regression analysis, as was the case for the four previously cited studies. One problem confronting these researchers and program evaluators is how to convey the meaning of a significant logistic coefficient in a manner that practitioners find meaningful and useful (Cabrera, 1994).

Various researchers have stressed the importance of not only reporting whether parameter estimates, such as logistic regression coefficients, are statistically significant but also

the need to provide information that can be used to judge the practical significance of these parameter estimates (Fisher, 1925; Cohen, 1969, 1988; Fraas & Newman, 2000; Levin & Robinson, 2000; Robinson & Levin, 1997; Thompson, 1996). Since a logistic regression coefficient measures the change in the log of the odds of the event occurring associated with a one-unit change in the independent variable, many practitioners may find such a value difficult to use in order to judge the practical significance of the change. Thus the question addressed in this paper is: What information regarding statistically significant logistic regression coefficients can researchers provide to practitioners that will allow them to best deal with the concept of practical significance?

This paper presents a method of expressing a logistic coefficient as a series of probability values where each value represents the change in a person's probability of belonging to the group assigned the value of one in the dependent variable (e.g., the teachers who were classified as belonging to the late-stages-of-concern group in the Graham [2002] study) associated with a one-unit change in the independent variable. We believe these probability values will provide researchers, program evaluators, and practitioners with information that will assist them in judging the practical significance of a change in the dependent variable that is associated with a given change in the independent variable as measured by the coefficient.

In addition to presenting this method of expressing a logistic coefficient as a series of probability values, this paper describes how a researcher or a practitioner can use a computer program, which is used in conjunction with the Microsoft® Excel software, to generate these probability values. A researcher or practitioner who wants to use the program can enter these commands (see Appendix A) into the Microsoft® Excel software or they can request a copy of the program from the authors by sending such a request to jfraas@ashland.edu.

Interpreting Logistic Regression Coefficients

In a logistic regression model, the relationship between the binary dependent variable and a given independent variable is assumed to follow a logistic function (Collett, 1991). The logistic function takes the following form:

$$\ln \frac{\text{Prob(event)}}{(1 - \text{Prob(event)})} = B_0 + B_1X_1 + B_2X_2 + \ldots + B_pX_p$$

where:

1. Prob(event) refers to the probability that the person will belong to the group assigned a value of one in the original dependent variable.
2. The symbols X_1, X_2, and X_p represent the independent variables.
3. The symbols B_0, B_1, B2, and B_p refer to the intercept and coefficients.

To estimate the logistic regression model, the logistic curve is fitted to the actual data. In the process of fitting the logistic curve to the data the values of the dependent variable are transformed to the logarithm of the odds (i.e., the log odds) that a person will belong to the group assigned a value of one in the original dependent variable. As noted by Cizek and Fitzgerald (1999), this transformation results in log odds values that are measured on an equal interval scale. Often these log odds values are referred to as logits, which is the contraction of the terms logistic and units.

The logistic transformation of the dependent variable causes the coefficients estimated for a logistic regression model to differ from those obtained from an Ordinary Least Squares (OLS) regression model. As noted by Cabrera (1994), "unlike OLS, the metric of the individual coefficients under logistic regression is expressed in terms of logits rather than in terms of the original scale of measurement" (p. 245). Thus the logistic coefficient for a given independent

variable is interpreted as the change in the log odds associated with a one-unit change in that variable.

To illustrate the interpretation of a logistic coefficient, consider the results of a study conducted by Graham (2002). In Graham's study, the dependent variable identified whether teachers were considered to be in the early or late stages of concern regarding the implementation of educational innovation in their schools. Teachers in the early-stages-of-concern group were assigned values of zero, while the teachers in the late-stages-of-concern group were assigned values of one. The logistics regression coefficient for the independent variable that represented the number of years the teachers worked with their current principals was statistically significant. The value for this variable, which was .161, indicates that an additional year of working with the current principal was associated with an increase of .161 in the log odds of belonging to the late-stages-of-concern group.

Practitioners, policy makers, and researchers may find that such an interpretation of a logistic regression coefficient does not provide information that readily lends itself to judging the practical significance of the coefficient. We believe it is important for researchers who use logistic regression to express the change in the dependent variable associated with a change in an independent variable in a manner that practitioners and policy makers can use to judge the educational or practical significance of the change.

Methods Commonly Used to Interpret Logistic Regression Coefficients

Two commonly used methods of interpreting statistical significance are (a) the odds ratio value, and (b) Delta-p statistic. With regards to the using of the odds ratio value, Cizek and Fitzgerald (1999) stated that: "Information related to the odds--as opposed to log odds--of an event occurring is easier to understand and communicate" (p. 230). Commenting on the use of

the Delta-p statistic in conjunction with a logistic regression coefficient, Petersen (1985) suggested that it is a suitable method of estimating the change in the dependent variable associated with the overall change in the independent variable.

The odds ratio value method. To illustrate how a researcher could communicate the meaning of a logistic regression coefficient by using the odds ratio value, consider the logistic regression coefficient for the variable representing the years of working with the principal in the Graham (2002) study, which was .161. Raising *e*, which is the base of the natural logarithm, to the power of .161, produces an odds ratio value of 1.17. This value indicates the factor by which the odds change when the years of working with the principal increases by one year. Thus an increase of one unit in the independent variable, which is one additional year of working with the principal, increases the odds of belonging to the late-stage-of-concern group (i.e., the group assigned a value of one in the dependent variable) by a factor of 1.17.

Delta-p statistic method. The Delta-p statistic estimates that change in the probability of a person belonging to the group assigned a value of one in the dependent variable that is associated with a one-unit change in a given independent variable. It is important to note that this change is calculated for only one initial probability value. The initial probability value used is set equal to the mean of the dependent variable, which represents the proportion of the sample belonging to the group assigned a value of one in the dependent variable. The calculation of the Delta-p value is based on the following equation:

$$P_c = \frac{e^{\ln(Pi/(1-Pi))+b}}{1+e^{\ln(Pi/(1-Pi)+b}} - Pi$$

where:

1. P_c represents the change in the probability value.
2. Pi represents the initial probability value. This value is set equal to the proportion of the sample belonging to the group assigned the value of one in the dependent variable (i.e., the sample mean of the dependent variable).
3. The symbol *e* represents the base of the natural logarithm.
4. The symbol *b* represents the logistic regression coefficient for the given predictor variable.

The interpretation of the Delta-p statistic, which is calculated as the initial probability equal to the mean of the dependent variable, depends on whether the nature of the independent variable is continuous or categorical. If the independent variable consists of continuous values, the Delta-p statistic indicates the change in the initial probability of belonging to the group assigned a value of one in the dependent variable associated with a one-unit increase in the independent variable. If the independent variable represents categories, the Delta-p statistic indicates the change in the initial probability of belonging to the group assigned a value of one in the dependent variable if the person is a member of the group assigned a value of one rather than a member of the group assigned a value of zero in the independent variable.

To illustrate the calculation and interpretation of the Delta-p statistic, consider the results produced by the Graham (2002) study in which the mean of the dependent variable was .41 and the logistic coefficient for the continuous independent variable of years working with the principal was .161. The Delta-p value would be calculated as follows:

$$P_c = \frac{e^{\ln(.41/(1-.41))+.161}}{1+e^{\ln(.41/(1-.41))+.161}} - .41$$

$P_c = .45 - .41$

$P_c = .04$

This value indicates that for an additional year of working with the principal the logistic model predicts an increase of .04 in the initial probability of belonging to the late-stage-of-concern group, which was .41.

A Delta-p statistic can also be calculated for an independent variable that consists of categories. To illustrate the calculation and interpretation of the Delta-p value for a categorical independent variable, consider once again the results produced by the Graham (2002) study, in which the categorical independent variable of tenure used values of zero and one to represent teachers without and with tenure, respectively. Since the logistic coefficient for this tenure variable was -.32, the Delta-p value would be calculated as follows:

$$P_c = \frac{e^{\ln(.41/(1-.41))-.32}}{1+e^{\ln(.41/(1-.41))-.32}} - .41$$

$P_c = .34 - .41$

$P_c = -.07$

This value indicates that for the teachers with tenure, the logistic model predicts a decrease of .07 in the initial probability of belonging to the late-stage-of-concern group when compared to teachers without tenure.

As previously demonstrated, researchers can calculate a Delta-p statistic for both continuous and categorical independent variables. It should be noted, however, that since there are no known procedures to estimate the statistical significance of a Delta-p statistic, St. John (1991) and Cabrera (1994) recommend that Delta-p statistics be calculated and interpreted only for statistically significant logistics coefficients. It should be mentioned that the logistic

regression coefficient for the tenure variable in the Graham (2002) study was not statistically significant. The Delta-p statistic was calculated for the tenure variable only for the purpose of providing an illustration of how such a value is calculated and interpreted for a categorical independent variable.

A Method Designed to Generate a Series of Changes in the Initial Probability Values

Although both the odds ratio value and the Delta-p statistic methods of reporting and interpreting statistically significant logistic regression coefficients provide information that is somewhat more useful than that provided by the log odds values, we believe that even more useful information can be provided to practitioners by expanding on the concept incorporated into the Delta-p statistic method. We believe the best way to report coefficients in order for researchers and practitioners to deal with practical significance is in the form of the change in the probability of belonging to the group assigned one in the dependent variable, which is the basis of the Delta-p statistic method. However, unlike the Delta-p statistic method, we believe the change in the initial probability should be calculated for a range of initial probabilities rather than just one such value.

The reason we believe the concept of the Delta-p statistic needs to be expanded centers on that fact that in logistic regression the actual data are not linearly related to the independent variable even though the log-transformed data are linearly related to it. Thus the change in the probability of belonging to the group assigned a value of one in the dependent variable is not constant for various initial probability values. Since the Delta-p statistic is calculated for only one initial probability value, it does not allow a practitioner to assess the degree of variability of the changes in the initial probability associated with a specified change in the independent variable for a relevant range of initial probability values. That is, a practitioner may find that the

changes in the initial probabilities are practically significant for those people with certain initial probabilities but not for those with other initial probability levels.

We believe the method that provides the most useful information in terms of allowing practitioners to judge the practical significance of a logistic regression coefficient is one that expands the Delta-p statistic method by including various initial probability values. The method we are proposing, along with our computer program, calculates the change in the probability value of belonging to the group assigned a value of one in the dependent variable for a specified change in an independent variable for a range of initial probability values. We believe that the changes in the initial probability values calculated through this method will provide practitioners a more complete picture regarding these changes, which will allow them to better deal with the issue of practical significance.

Values Generated by the Proposed Method

The method we are proposing produces one key series of probability values and three important individual probability values: (a) a series of changes in the probability of the event occurring for a given change in the independent variable that corresponds to a series of initial probability values, (b) the Delta-p statistic, (c) the initial probability value at which the absolute value of the change in the probability of the event occurring is maximized for a given change in the independent variable, and (d) the change in the probability of the event occurring for a given change in the independent variable for which the absolute value of this change is maximized. We believe using these values will better enable practitioners to accomplish the task of gauging the practical significance of the logistic regression coefficient.

A series of changes in the probability of an event occurring. A series of probability values are calculated in which each value measures the change in the probability of belonging to

the group assigned a value of one in the dependent variable for a given change in an independent variable at a specified initial probability level. One issue that must be addressed regarding the calculation of these changes in the initial probability values is: What initial probability values should be used? We recommend that the range of the initial probabilities be determined by the minimum and maximum probabilities predicted by the logistic regression model for the sample.

Once the minimum and maximum initial probability values are determined, an interval must be specified to generate a series of initial probabilities. We suggest that the use of a .05 interval generally will provide enough information to assess the practical significance of the coefficient. Although the program is designed to handle intervals as small as .01, we believe one should generally use an interval of .05 or .10. The number of probability values generated by intervals smaller than .05 can be overwhelming, while the number of probability values generated by intervals larger than .10 may provide too little information.

After the series of initial probabilities has been generated, the final probability (P_f) of belonging to the group assigned a value of one in the dependent variable for a specified change in the independent variable, is calculated for each initial probability by using the following equation:

$$P_f = \frac{e^{\ln(Pi/(1-Pi))+b}}{1+e^{\ln(Pi/(1-Pi))+b}}$$

Once these final probability values are calculated, a series of corresponding changes in the probability values is calculated by subtracting each initial probability from its initial final probability.

As is the case for the Delta-p statistic, the interpretation of these changes in the probability values depends on whether the independent variable is continuous or categorical. If

the independent variable consists of continuous values, these values indicate the changes in the probability of belonging to the group assigned a value of one in the dependent variable associated with the specified increase in the independent variable. If the independent variable represents categories, these values indicate the change in the initial probability of belonging to the group assigned a value of one in the dependent variable if the person is a member of the group assigned a value of one rather than a member of the group assigned a value of zero in the independent variable.

An examination of these changes in the probabilities that the event will occur will assist practitioners in judging the practical significance of the logistic regression coefficient in two ways. First, the practitioners can readily gauge the change in the dependent variable associated with a given change in the independent variable in meaningful terms (i.e., the change in the probability of the event occurring). The practitioners can judge whether an increase of, say, .10 in the probability that the event will occur for a given change in the independent variable is large enough to have educational importance (i.e., it is practically significant). Second, an examination of the series of changes in the probability values will reveal the degree of variability in these values across the range of initial probability values. Such an examination will allow the practitioners to determine whether the changes in the probability that the event will occur is practically significant across the entire range of initial probabilities or only for a specific subgroup of initial probability values.

Individual probability values. In addition to the series of changes in the probability of the event occurring for a given change in the independent variable, practitioners may find three individual values useful in judging the practical significance of a logistic regression coefficient. The first of these values is the Delta-p statistic. As previously described, the Delta-p statistic

indicates the change in the initial probability value, which is set equal to the mean of the dependent variable, associated with a given change in the independent variable. This Delta-p statistic may be useful to practitioners in judging the practical significance of a logistic regression coefficient when the variability of the changes in the probability of the event occurring is minimal across the relevant range of initial probability values. In such a case, the Delta-p statistic, which is one value, may suffice as a means of judging the practical significance of the coefficient.

The second individual value is the initial probability at which the largest absolute change in the probability associated with a specified change in the independent variable occurred. To identify this initial probability value, the numerator of the first derivative with respect to the initial probability of the following expression is set equal to zero:

$$\frac{e^{\ln(Pi/(1-Pi))+b}}{1+e^{\ln(Pi/(1-Pi))+b}}$$

This procedure produced the following quadratic expression:

$$\frac{1\pm\sqrt{e^{b}}}{1-e^{b}}$$

The initial probability value at which the absolute change in the probability is maximized is calculated by substituting the value of the logistic regression coefficient into the following portion of the quadratic expression:

$$\frac{1-\sqrt{e^{b}}}{1-e^{b}}$$

The third individual value is the actual value (i.e., the value with the sign included) for this maximized absolute value of the change in the initial probability.

The actual value of the maximum absolute value of the change in the probability and its corresponding initial probability may provide useful information to practitioners in two ways. First, if the actual value of the largest absolute change in the initial probability does not reach the level deemed practically significant by the practitioners, they will quickly know that practical significance is not reached at any point within the range of relevant initial probability values. Second, practitioners can identify whether the initial probability at which the absolute value of the change is maximized is located at a point on the initial probability scale that is common or uncommon for the subjects included in the study.

Computer Program Used to Generate Key Values

The computer program we designed, which is used in conjunction with the Microsoft® Excel software, calculated and/or listed the following: (a) a series of initial probability values, (b) a series of corresponding final probability values associated with a given change in the independent variable, (c) a series of corresponding differences between the final and initial probability values, (d) the Delta-p statistic, (e) the initial probability value at which the change in the absolute value of the initial probability is maximized for a given change in the independent variable, and (f) the actual value of the maximum absolute change in the initial probability for a given change in the independent variable. In addition to generating these values, the program is designed to graph the curve depicting the relationship between the initial probability values (X axis) and the corresponding changes in the probability values (Y axis) for a given change in the independent variable. This visual presentation may assist the researcher in assessing the degree of change in the probability levels throughout the range of initial probability values.

A description of the input values used in the program. The researcher must identity six key input values, which are entered into the first page of the computer program. To assist in

describing these values and how they are entered into the program, a copy of the first page that contains the six key values used in the Graham (2002) study is presented in Figure 1.

Insert Figure 1 about here

The first two values entered into the program are: (a) the size of the change in the independent variable specified by the researcher and (b) the statistically significant logistic regression coefficient being evaluated. These values are entered into the rows entitled "Change in the Independent Variable" and "Coefficient Value," respectively. In the Graham (2002) study values of 1 and .161 were entered for the change in the independent variable and the logistic regression coefficient, respectively (see Figure 1).

The third value entered into the program is the proportion of subjects in the sample who belonged to the group assigned a value of one in the dependent variable. This value, which is equal to the mean of the dependent variable, is entered into the row entitled "Proportion of Cases with a Value of 1." The proportion of subjects in the sample used in the Graham (2002) study was .41 (see Figure 1).

Since we are interested in generating a series of changes in the probability of the event occurring for a corresponding series of initial probability values, the researcher must identify a relevant range for these values. Thus the fourth and fifth numbers entered into the program are the minimum and maximum values of the initial probability values. These values are entered into the rows entitled "Minimum Value" and "Maximum Value." We suggest the minimum and maximum values be set approximately equal to the minimum and maximum probability values predicted by the logistic regression model for the subjects contained in the sample. The

approximate minimum and maximum probability values for the Graham (2002) study were .20 and .70, respectively (see Figure 1).

The sixth value entered into the program is an interval that will be used to generate a series of initial probability values located between the minimum and maximum initial probabilities values, which were previously identified. The specified interval is placed in the row entitled "Interval." The interval was set at .05 in the Graham (2002) study (see Figure 1).

A description of the output values produced by the program. Once the six input values are entered into the first page of the program, three series of probability values and three individual probability values are generated and listed on the second page of the program. To assist in describing these values, a copy of the second page of the program for the Graham (2002) study is contained in Figure 2.

Insert Figure 2 about here

The first of the three series of probability values consists of the initial probability values, which are listed in the program under the heading "Initial P Values." These values begin with the minimum initial probability value stipulated on the first page of the program and increase by the specified interval until the stipulated maximum initial probability is reached. In the Graham (2002) study these values ranged from the minimum initial probability value of .20 and increased by intervals of .05 until the maximum probability value of .70 was reached (see Figure 2).

The second series of values contains the final probability values, which are calculated from the initial probability values. These final probability values, which indicate the probability of the event occurring once the independent variable was increased by the specified amount, are

listed in the program under the heading "Final P Values." Eleven final probability values were calculated for the Graham (2002) study (see Figure 2).

The third series of numbers consists of the differences between the final and initial probability values. These values, which represent the changes in the initial probability values associated with the specified change in the independent variable, are listed in the program under the heading "Change in P Values." These changes in the initial probability values ranged from a low of .027 to a high of .040 for the Graham (2002) study (see Figure 2).

In addition to these three series of numbers, the second page of the program contains three individual probability values. The first of these values is the Delta-p statistic. This value is listed in the row entitled "Delta-p Statistic for '*p*'," where the mean of the dependent variable is actually printed rather than the symbol *p*. As previously described, the Delta-p statistic indicates the change in the initial probability value, which is set equal to the mean of the dependent variable, associated with a specified change in the independent variable. The Delta-p statistic for the Graham (2002) study was .040 (see Figure 2).

The second of the three individual values listed on the second page of the program is the initial probability at which the largest change in the initial probability occurs for the specified change in the independent variable. This value is entered in the row entitled "P value at which the absolute change in P is the largest for the specified change in the independent variable." In the Graham (2002) study, the change in the initial probability value at which the absolute change was the largest for the specified change of one year in the years working with the principal was .480 (see Figure 2).

The last of the three individual values listed on the second page of the program is the actual value of the maximum absolute change in the initial probability for the specified change in

the independent variable. This value, which occurs at the initial probability previously identified by the program, is listed in the row entitled "Largest absolute change in P expressed as P." In the Graham (2002) study the maximum change in the probability of a teacher belonging to the late-stage-of-concern group associated with a one-year change in the number of years working with the principal was .040, which, as previously noted, occurred at the initial probability of .480 (see Figure 2). It should be noted that although this maximum value of .040 appears to equal the Delta-p statistic, which was recorded for a different initial probability value, it does exceed the Delta-p statistic if a sufficient number of decimal places were displayed.

The third page of the program contains a graph of the changes in the probability values (Y axis) with their corresponding initial probability levels (X axis). The purpose of the graph is to provide the researcher a visual depiction of this relationship. The graph generated for the Graham (2002) study is presented in Figure 3.

Insert Figure 3 about here

Assessing the Graham Study Information

The statistically significant logistic regression for the independent variable representing the number of years the teachers worked with the current principals, which was .161, indicated that a one-year increase was associated with a .161 log odds change in belonging to the late-stage-of-concern group. We believe that most practitioners would find that such an interpretation provides little useful information with respect to judging the practical significance of the coefficient. Although somewhat more useful, we believe that practitioners would also find the corresponding odds value of 1.17 difficult to use when judging the practical significance of

the coefficient. We believe that practitioners will find the change in probability values corresponding to a specified change in the independent variable, which are listed on the second page of our computer program do assist them in judging the practical significance of the coefficient.

A review of the values listed on the second page of the computer program for the Graham (2002) study reveals that the changes in the initial probability values for a one-year change in the teachers' years of working with their principal ranged from a minimum of .027 at the initial probability level of .20; reached a maximum of .04 at the initial probability of .48; and declined in size to .033 at the initial probability of .70 (see Figure 2). The graph presented on the third page of the program provides a visual aid depicting this relationship between the changes in the initial probability values and their corresponding initial probability values (see Figure 3).

After reviewing these values, two conclusions can be made. First, the differences among the changes in the initial probability values associated with a one-year change were not substantial across the range of initial probability values. Thus the level of practical significance of the logistic regression coefficient can be accurately judged by using the single value of the Delta-p statistic, which was .04. Second, using the Delta-p value, it was concluded that the educational significance of the change in moving teachers from the early-stage-of-concern group to the late-stage-of-concern group was minimal. That is, the change in the probability is estimated to be approximately .04.

One very important point must be considered, however, before a final conclusion is made regarding the practical significance of this statistically significant logistic regression coefficient. That is, before practitioners judge the practical significance of changes in the probability of belonging to the group assigned a value of one in the dependent variable, they may want to

evaluate the changes associated with multiple unit changes in the independent variable (Norušis, 1999, pp. 39-40). This point is demonstrated by the changes in the probability levels corresponding to changes in the number of years of working with the principal that exceed one year (Graham, 2002). To illustrate this point, consider the differences in the changes in the initial probability levels of teachers who have four additional years of experience working with their principals, which is approximately one standard deviation unit of this independent variable. To obtain these probability values, the researcher would specify the change in the independent variable as 4 rather than one and enter this value next the title "Change in the Independent Variable" in the computer program.

The series of changes in the probability values generated by the program, which are listed in Figure 4, indicate that the changes in the initial probability values for teachers with four additional years working with their principals compared to teachers without the additional four years range from a minimum of .123 at the initial probability level of .20; reached a maximum of .16 at the initial probability of .42; and declined in size to .116 at the initial probability of .70. These values suggest that although the increase in the probability that teachers would be reclassified from the early-stage-of-concern group to the late-stage-of-concern group associated with a one-year change in the years working with the principal may be judged to have minimal educational importance, difference between teachers with little and significant work experience with the principal may be judged to have educational importance.

Insert Figure 4 about here

Summary

Program evaluators and educational researchers who use logistic regression to analyze dichotomous dependent variables are faced with the task of assisting practitioners in judging whether a statistically significant coefficient is also practically significant. The method and the computer program presented in this paper are designed to provide program evaluators and researchers with values that will assist them in assessing the practical significance of such a coefficient. The program requires the researcher to enter six values: (a) a specified change in the independent variable, (b) the logistic coefficient value, (c) the proportion of the subjects in the sample assigned a value of one in the dependent variable, (d) a minimum initial probability value, (e) a maximum initial probability value, and (f) an interval used to generate a series of values between the minimum and maximum initial probability values.

Once these six values are entered into the program, it generates a series of changes in the initial probabilities associated with a specified change in the independent variable. In addition to this series of values, the program calculates three individual values. The first of these values is the Delta-p statistic. This value indicates the change in the initial probability value, which is set equal to the mean of the dependent variable, associated with a specified change in the independent variable. The second individual value is the initial probability at which the largest absolute change in the initial probability occurs. The third individual value represents the actual value of the maximum absolute value of the change in the initial probability for a specified change in the independent variable. In addition to these values, the program also provides a graph that depicts the relationship between the changes in the probability values (Y axis) and their corresponding initial probability levels (X axis).

We believe that the values generated by the method and the computer program described in this paper will provide researchers, program evaluators, and practitioners with information that will assist them in judging the practical significance of a logistic regression coefficient in a number of ways. First, practitioners should find the changes in probability values to be easier to use when judging the practical significance of the coefficient as compare to the odds ratio value (i.e., the logistic regression coefficient) or even the odds ratio value. Second, the series of changes in the initial probability values associated with a specified change in the independent variable will allow practitioners to judge whether the Delta-p value provides enough information to judge the practical significance of the coefficient. Third, if a practitioner believes the practical significance of the coefficient should be judged in terms of changes in the independent variable that are greater than one unit, the program can easily generate the essential values that reflect such a change.

The need to address the practical significance of statistically significant findings is an important task for researchers, program evaluators, and practitioners to undertake. We believe the implementation of the method through the use of the computer program, as described in this paper, will assist researchers, program evaluators, and practitioners in the task of accessing the practical significance of logistic regression coefficients.

References

Cabrera, A.F. (1994). Logistic regression analysis in higher education: An applied perspective. In J.C. Smart (Ed.), *Higher education: Handbook of theory and research* (pp. 225-254). New York: Agathon Press.

Cizek, G.H. & Fitzgerald, S.M. (1999, January). An introduction to logistic regression. *Measurement and Evaluation in Counseling and Development, 31*(4), 223-245.

Collett, D. (1991). *Modeling binary data.* New York: Chapman Hall.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Fisher, R.A. (1925). *Statistical methods for research workers.* London: Oliver & Boyd.

Fraas, J.W. & Newman, I. (2000, October). *Testing for statistical and practical significance: A suggested technique using a randomization test.* Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL.

Graham, J. (2002). *Teacher acceptance of change.* Unpublished doctoral dissertation, Ashland University, Ashland, OH.

Hendel, D.D. (2001, April). *The relative contribution of participating in a first-year seminar on student satisfaction and retention into the sophomore year.* Paper presented at the annual meeting of the American Educational Association, Seattle, WA.

Levin, J.R. & Robinson, D.H. (2000). Rejoinder: Statistical hypothesis testing, effect-size estimate, and the conclusion coherence of primary research studies. *Educational Researcher, 29*(1), 34-36.

McCoach, D.B. & Siegle, D. (2001, April). *Why try? Factors that differentiate underachieving gifted students from high achieving gifted students.* Paper presented at the annual meeting of the American Educational Association, Seattle, WA.

Norušis, M.J. (1999). *SPSS regression models™ 10.0.* Chicago, IL: SPSS, Inc.

Petersen, T. (1985). A comment on presenting results from logit and probit models. *American Sociological Review, 50*(1), 130-131.

Robinson, D.H. & Levin, J.R. (1997). Reflections on statistical and substantive significance, with a slice of replication. *Educational Researcher, 26*(5), 21-26.

Schreiber, J.B. (2002, Spring). Scoring above the international average: A logistic regression model of the TIMSS advanced mathematics exam. Multiple *Linear Regression Viewpoints, 28*(1), 22-30.

St. John, E.P. (1991). What really influences minority attendance? Sequential analyses of the high school and beyond sophomore cohort. *Research in Higher Education, 32*(2), 141-158.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher, 25*(2), 26-30.

Appendix A

Interested readers can obtain the worksheets used to calculate the various probability estimates for a given logistic regression coefficient as discussed in the previous sections in either of two ways. They can design the required worksheets by utilizing the Microsoft® Excel computer software in conjunction with the commands outlined in this section or they can request copies of the program from the authors. Send such requests to the following e-mail address: jfraas@ashland.edu.

The commands used in conjunction with the Microsoft® Excel computer software, which are presented in this section, are divided into four sections. The first set of commands is designed to label the three worksheets included in the spreadsheet. These commands, which are printed in bold type, are as follows:

1. Double-click on the tab for **Sheet 1**.
2. Enter the name **Parameters**.
3. Double-click on the tab for **Sheet 2**.
4. Enter the name **Calculations**.
5. Double-click on the tab for **Sheet 3**.
6. Enter the name **Chart**.

These commands name the three worksheets Parameters, Calculations, and Chart.

The second set of commands is used to construct the Parameters worksheet. These commands are as follows:

1. Click on the tab titled **Parameters**
2. Enter in the specified cells the following text:

   (a) Cell A1 -- **The Calculation of the Changes in Initial Probability**

(b) Cell A2 -- **Values for Logistic Regression Coefficients**

(c) Cell A3 -- **Change in the Independent Variable:**

(d) Cell A4 -- **Coefficient Value:**

(e) Cell A5 -- **Proportion of Cases with a Value of 1:**

(f) Cell A6 -- **Minimum Value:**

(g) Cell A7 -- **Maximum Value:**

(h) Cell A8 -- **Interval:**

(i) Cell C3 -- **[the size of change in the independent variable]**

(j) Cell C4 -- **[the logistic regression coefficient for the variable]**

(k) Cell C5 -- **[average of the dependent variable]**

(l) Cell C6 -- **[the lowest predicted probability for the sample data]**

(m) Cell C7 -- **[the highest predicted probability for the sample data]**

(n) Cell C8 -- **[interval size between the min. and max. values]**

3. Change the background of the range **B3:B8** to indicate the location of the data to be entered.

The third set of commands is used to construct the Calculation worksheet. These commands are as follows:

1. Click on the tab titled **Calculation**.
2. Enter in the specified cells the following text:

(a) Cell A1 -- **Initial P Values**

(b) Cell B1 -- **Final P Values**

(c) Cell C1 -- **Change in P Values**

(d) Cell D4 -- **Delta-P Statistic for**

(e) Cell D7 -- **P value at which the absolute change in P**

(f) Cell D8 -- **is the largest for the specified change in the**

(g) Cell D9 -- **independent variable**

(j) Cell D11 -- **Largest absolute change in P expressed as P**

3. Enter in the specified cells the following formula:

(a) Cell A2 -- **=Parameters!B6**

(b) Cell A3 -- **=IF(A2=" "," ",IF(A2+Parameters!$B$8<**

(Parameters!$B$7+0.0001) ,A2+Parameters!$B$8," "))

(c) Cell B2 -- **=A2+C2**

(d) Cell B3 -- **=IF(A3=" "," ",A3+C3)**

(e) Cell C2 -- **=IF(A2=" "," ",RAND()*0+(EXP((LN(A2/(1-A2))**

+(Parameters!$B$4*Parameters!$B$3))))/(1+EXP((LN(A2/(1-A2))

+(Parameters!$B$4*Parameters!$B$3))))-A2)

(f) Cell C3 -- **=IF(A3=" "," ",(EXP((LN(A3/(1-A3))**

+(Parameters!$B$4*Parameters!$B$3))))/(1+EXP((LN(A3/(1-A3))

+(Parameters!$B$4*Parameters!$B$3))))-A3)

(g) Cell F4 -- **=(EXP((LN(F2/(1-F2))+(Parameters!$B$4*Parameters!$B$3))))/**

(1+EXP((LN(F2/(1-F2))+(Parameters!$B$4*Parameters!$B$3))))-F2

(h) Cell F9 -- **=(1-SQRT(EXP((Parameters!B4*Parameters!$B$3))))/(1-**

(EXP((Parameters!B4*Parameters!$B$3))))

(j) Cell F11 -- **=(EXP((LN(F9/(1-F9))+(Parameters!$B$4*Parameters!$B$3))))/**

(1+EXP((LN(F9/(1-F9))+(Parameters!$B$4*Parameters!$B$3))))-F9

4. Copy the range **A3:C3** and paste it into the range **A4:C101**

5. Insert the following names:

   (a) "change_in_p" refers to **=OFFSET('The Calculation'!$B$2,0,0,COUNT ('The Calculation'!$B$2:$B$101),1)**

   (b) "interval" refers to **=OFFSET('The Calculation'!$A$2,0,0,COUNTA ('The Calculation'!$A$2:$A$101),1)**

6. Save the workbook with the name 'Fraas'

The fourth set of commands is used to construct the Chart worksheet. These commands are as follows:

1. Click on the tab entitled **The Calculation**
2. Turn off the legend by clicking on the **Show Legend** box
3. Click on **Next>**
4. Click on the button for '**As object in:**'
5. Enter **'The Chart'**
6. Click on the tab titled **The Chart**
7. Move cursor to the body of the chart and right click
8. Click on **'Source Data...'**
9. In the Values: box enter **=Fraas!change_in_p**
10. In the Category (X) axis labels box enter **=Fraas!interval**
11. Click on **OK**
12. Save the Workbook

The Calculation of the Changes in Initial Probability Values for a Logistic Regression Coefficient

Change in the Independent Variable:	1	[the size of change in the independent variable]
Coefficient Value:	0.161	[the logistic regression coefficient for the variable]
Proportion of Cases with a Value of 1:	0.41	[average of the dependent variable]
Minimum Value:	0.2	[the lowest predicted probability for the sample data]
Maximum Value:	0.7	[the highest predicted probability for the sample data]
Interval:	0.05	[interval size between the min. and max. values]

Figure 1

Page 1 of the Computer Program that Contains the Results of the Graham Study with the Change in the Independent Variable Set at One Year

Initial P Values	Final P Values	Change in P Values
0.20	0.227	0.027
0.25	0.281	0.031
0.30	0.335	0.035
0.35	0.387	0.037
0.40	0.439	0.039
0.45	0.490	0.040
0.50	0.540	0.040
0.55	0.589	0.039
0.60	0.638	0.038
0.65	0.686	0.036
0.70	0.733	0.033

Delta-p Statistic for 0.41	0.039

P value at which the absolute change in P is the largest for the specified change in the independent variable	0.480

Largest absolute change in P expressed as P	0.040

Figure 2

Page 2 of the Computer Program that Contains the Results of the Graham Study with the Change in the Independent Variable Set at One Year

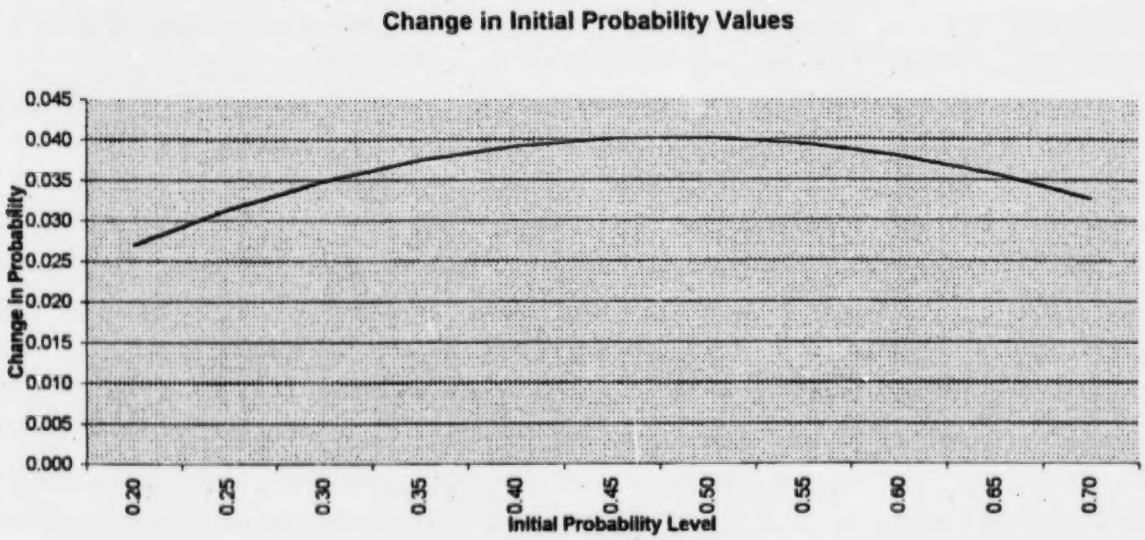

Figure 3

Page 3 of the Computer Program that Contains the Results of the Graham Study with the Change in the Independent Variable Set at One Year

Initial P Values	Final P Values	Change in P Values		
0.20	0.323	0.123		
0.25	0.388	0.138		
0.30	0.449	0.149	Delta-p Statistic for **0.41**	**0.160**
0.35	0.506	0.156		
0.40	0.559	0.159		
0.45	0.609	0.159	P value at which the absolute change in P	
0.50	0.656	0.156	is the largest for the specified change in the	
0.55	0.699	0.149	independent variable	**0.420**
0.60	0.741	0.141		
0.65	0.780	0.130	Largest absolute change in P expressed as P	**0.160**
0.70	0.816	0.116		

Figure 4

Page 1 of the Computer Program that Contains the Results of the Graham Study with the Change in the Independent Variable Set at Four Years

BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC

TM034549

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Expressing Logistic Regression Coefficients as Changes in Initial Probability Values: Useful Information for Practitioners	
Author(s): John W. Fraas; J. Michael Drushal; Jeff Graham	
Corporate Source: Ashland University, Ashland, Ohio	Publication Date: 10/17/02

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

X

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: John W Fraas	Printed Name/Position/Title: John Fraas, Trustees' Professor	
Organization/Address: Ashland University, 401 College Avenue, Ashland, Ohio, 44805	Telephone: 419-289-5930	FAX: 419-289-5989
	E-Mail Address: jfraas@ashland.edu	Date: 10/21/02

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: info@ericfac.piccard.csc.com
WWW: http://ericfacility.org

EFF-088 (Rev. 9/97)
PREVIOUS VERSIONS OF THIS FORM ARE OBSOLETE.